What time travelers cannot *not* do (but are responsible for anyway)

Joshua Spencer

Abstract

The Principle of Alternative Possibilities is the intuitive idea that someone is morally responsible for an action only if she could have done otherwise. Harry Frankfurt has famously presented putative counterexamples to this intuitive principle. In this paper, I formulate a simple version of the Principle of Alternative Possibilities that invokes a course-grained notion of actions. After warming up with a Frankfurt-Style Counterexample to this principle, I introduce a new kind of counterexample based on the possibility of time travel. At the end of the paper, I formulate a more sophisticated version of the Principle of Alternative Possibilities that invokes a certain fine grained notion of actions. I then explain how this new kind of counterexample can be augmented to show that even the more sophisticated principle is false.

Key Words: Time Travel; Principle of Alternative Possibilities; Frankfurt

0. Introduction

The idea that someone is morally responsible for an action only if she could have done otherwise is intuitively appealing. It seems perfectly acceptable, for example, to defend our actions by pointing out that we could not do otherwise. If I have been attacked and both of my legs have been broken, then it seems illegitimate to criticize me for failing to run away; I could not have done otherwise. If all of my bodily movements are being controlled by outside alien

forces, then it is illegitimate to criticize me for driving my car into a neighbor boy; I could not have done otherwise.

Harry Frankfurt has famously argued that this intuitively appealing idea is mistaken. In the literature, a vast field has grown in response to Frankfurt's cases. In this paper, I formulate a simple principle which is meant to capture the idea that someone is morally responsible for an action only if she could have done otherwise. After warming up with a Frankfurt-Style Counterexample to this principle, I introduce a new kind of counterexample based on the possibility of time travel. This new kind of counterexample avoids some unnecessary complications that come along with Frankfurt-Style Counterexamples. Moreover, this new kind of counterexample is strong enough to overcome even more sophisticated principles that are inspired by the intuitive idea. For example, at the end of the paper I explain how this new kind of counterexample can be extended to show that even a sophisticated principle of alternative possibilities, a principle that invokes a certain fine grained notion of actions, is false.

1. Frankfurt-Style Counterexamples

The idea that someone is morally responsible for an action only if she could have done otherwise can be spelled out in a very straightforward way. Let's call this straightforward formulation 'The Principle of Alternative Possibilities' or 'PAP':

(PAP) Necessarily: for any person S and any action A, S is morally responsible for performing A only if there is some action A* such that S could have done A* while failing to do A.

Two important points should be made about the variables in (PAP). First, in this principle, the variables 'A' and 'A*' need not range over fully determinate actions.[1] For example, the act of driving one's car into a neighbor boy is an acceptable substitution instance for these variables. This is true in spite of the fact that there are several different ways that one might drive one's car into a neighbor boy. One might, for example, drive into him while backing out of a parking space or one might drive into him while speeding down the road. These ways of driving into a neighbor boy might, themselves, be done in further (more specific) ways. Of course, there are fully determinate ways of doing actions, ways of doing actions which cannot be more specific. Moreover, these fully determinate actions might play an important role in moral philosophy and we will return to consider them later in this essay. But, for now, let's focus on (PAP) and the, perhaps naïve, sense of action according to which an action need not be fully determinate.

A second important point is that refrainings are acceptable substitution instances of 'A' and 'A*'. So, for example, one way I can do otherwise than run over the neighbor boy is by refraining from doing so. Whether refrainings are technically actions is open to debate. However, I do not wish to enter that debate. Let's simply take the variables in (PAP) to be wide enough to range over both actions and refrainings.

[1] Some people make the distinction between act types and act tokens. They might say, for example, that the act of returning a book to the library is an act type whereas the particular way that I returned *An Essay on Free Will* this week is an act token. It seems to me that the relationship between returning a book and returning a book in a particular way is more like the relationship of a determinable to a determinate rather than a type to a token. But, nothing of importance in this essay will hinge which relationship obtains between these two actions.

Many people have thought that Frankfurt-Style Counterexamples show this principle to be false.[2] Here is one of those famous counterexamples: Jones wants Smith to be dead and is deliberating about whether to kill him. Black also has a strong desire to see Smith perish but Black prefers to avoid killing Smith himself. It turns out that Black is a skilled neurosurgeon and he comes up with a plan to guarantee Smith's death. He surreptitiously implants a device into Jones's brain which monitors Jones's decision making process. The device is made to detect whether or not Jones will go through with the killing of Smith. If the device were to detect that Jones will not go through with the killing, then a mind control mechanism would be activated, which would take over Jones's decision making process and thereby set Jones on a path toward carrying out the killing anyway. On the other hand, if the device were to detect that Jones will carry through with the killing, then the mind control mechanism would remain inert. After some period of deliberation, Jones decides to kill Smith and the mind control mechanism remains inert. In this situation, it seems that Jones is morally responsible for killing Smith. However, since the mechanism would have forced him to carry out the killing if he had been on the verge of refraining from doing so, Jones could not have done otherwise than kill Smith. That is, Jones could not have done some other action while refraining from killing Smith. This seems to show that (PAP) is false.

One might object to Frankfurt-Style Counterexamples by pointing out that devices, such as the one implanted in Jones's brain, can malfunction. Moreover, one might claim that if Jones were to have succeeded in doing otherwise than kill Smith, then the device would have had to have malfunctioned. But, I hope everyone will agree with me in assessing this complaint to be off the mark. Consider, for comparison, a prisoner who is being held in a high tech facility and

[2] Harry Frankfurt's famous counterexamples were first introduced in his (1969).

who cannot escape unless the locks on his cell malfunction. Although the locks on such a prisoner's cell might have malfunctioned, and, moreover, although those locks would have had to have malfunctioned, if the prisoner were to have succeeded in escaping, I think it is clear that these distant possibilities cannot ground the claim that the prisoner can escape. As long as the locks are well constructed, and as long as a malfunction is highly unlikely or physically impossible, then in all the scenarios accessible by the prisoner, he remains trapped in his cell. Similarly, although the device in Jones's head might have malfunctioned and, moreover, it would have had to have malfunctioned if Jones were to have succeeded in doing otherwise than kill Smith, the claim that Jones could have done otherwise than kill Smith cannot be grounded in such a distant possibility. It seems clear that as long as the device in his head is well constructed, as long as a malfunction is highly unlikely or perhaps physically impossible, then in all the scenarios accessible by Jones, he kills Smith. This point will be important to remember in the discussion that follows: distant possibilities do not act as guardian angels to the truth of claims about what we *can* do.

A second and far more serious objection involves claiming that Jones in fact could have done otherwise than kill Smith even though nothing he could have done would have allowed for Smith's survival. One might think that it is conceptually necessary that events are actions only if they are not compelled by certain outside forces. In the counterfactual circumstance, since Jones's bodily movements are compelled by mental events that were manipulated by the mechanism in his brain, those bodily movements do not constitute an act of killing. In the counterfactual circumstance, Jones refrains from performing a killing by doing something that led to the activation of the mechanism which, in turn, forced him to move in a way that resulted

in Smith's death.[3] So, even though Jones could not avoid bodily movements that result in Smith's death, he could have done otherwise than *kill* Smith by ensuring that those bodily movements do not constitute an *act* of killing. I take this to be a serious objection to Frankfurt-Style Counterexamples. However, in the next section, I will argue that even if this objection is sound, there is still another, new kind of counterexample to (PAP).

2. A Time Travel Counterexample to (PAP)

Our current physical theories are consistent with the possibility of time travel. Some models of General Relativity contain paths in space-time that an ordinary object can traverse at sub-luminal speeds and that allow such an object to arrive at some point in its past. These paths are called "closed time-like curves".[4] But, whether or not space-time contains closed time-like curves depends on the distribution of matter and the resulting curvature of space-time. Any restriction on possible distributions of matter and, hence, on space-time structures that rules out the possibility of closed time-like curves would constitute an arbitrary cutoff in modal space. Since there are no arbitrary cutoffs in modal space, it seems that closed time-like curves are possible and, hence, that time travel is possible.[5]

Moreover, some physical theories suggest that we might be able to manipulate the distribution of matter in our universe and create closed time-like curves. If, for example, we could manipulate matter to create a stable worm hole, then we might be able to create closed

[3] This, I take it, is a very rough characterization of the objection put forward in (Maria Alvarez, 2009).

[4] Arntzenius and Maudlin (2002) have a nice discussion of the possibility of time travel.

[5] Philip Bricker (1991) has an extensive discussion of principles of plenitude of possible structures. Such principles seem to support the kind of argument I am alluding to here.

time-like curves that can be used to travel through time.[6] Since we shouldn't rule out possibilities implied by viable physical theories on philosophical grounds alone, it seems that we should take seriously the possibility of time travel and even the possibility of time machines. But, the possibility of time machines and time travel seems to spell trouble for (PAP).

Suppose that Martin is a time traveler; he travels in a machine that can transport him to various points in earth's history.[7] During one of his many trips, Martin rescued a man from plummeting to his death. Let's suppose that the man rescued was a high-wire walker who was working without a net in a very desolate area with no one else nearby. Martin arrived in his machine just in time to see the man fall from his wire, head first toward the ground. Luckily, Martin quickly found a button labeled "Emergency Safety Net Release". Martin pressed the button and an emergency safety net was deployed across the field underneath the high-wire walker. The walker landed safely in the net and walked away from the situation unscathed. If Martin had not pushed the button and released the emergency safety net, then the high-wire walker would have fallen to his death.

As it turns out, this high-wire walker was Martin's grandfather at a very young age. At the time of his fall, Martin's high-wire walking grandfather had not yet met Martin's grandmother and he had not done anything that would have preserved his gametes for posterity. So, since Martin's grandfather would have died without any progeny, it is clear that (1) if Martin had not released the emergency safety net, then Martin would have never existed. But, of course, it's not possible for Martin to *do* anything if he never exists. In particular, (2) it's not possible for

[6] Paul Davies (2001) presents a clear and accessible account of the physics behind artificial worm holes and time travel.

[7] Perhaps it can do so by traveling through certain artificial worm holes.

Martin to *do* otherwise than release the emergency safety net if he never exists. Therefore, (3) if Martin had not released the emergency safety net, then he would not have done otherwise than release the emergency safety net. One thing that's obvious is that (4) it is broadly possible that Martin does not release the emergency safety net. After all, Martin might not have existed, in which case, he certainly would not have released the emergency safety net. It turns out, though, that from (3) and (4) it follows that (5) it is not the case that if Martin had not released the safety net, then he would have done otherwise. That is, it follows that the conditional 'if Martin had not released the safety net, then he would have done otherwise' is false. But, it seems that (6) if Martin could have done otherwise than release the safety net, then if he had not released the safety net, he would have done otherwise than release the safety net. It follows that (7) Martin could not have done otherwise than release the safety net.[8] Since Martin is morally responsible

[8] This argument is counterfactually valid. Let 'O' stand for the proposition that Martin does otherwise than release the safety net, let 'E' stand for the proposition that Martin exists at some time or other, and let 'R' stand for the proposition that he does release the safety net. Finally, let '◊' stand for 'it is possible that…' and '$◊_m$'stand for 'Martin can make it such that…', and let '□→' stand for the counterfactual conditional. Now we can formulate a formal version of this argument as follows:

(1) (~R □→ ~E)

(2) □(~E → ~O)

So, (3) (~R □→ ~O) [from (1) and (2)]

(4) ◊~R

So, (5) ~(~R □→ O) [from (3) and (4)]

(6) $◊_m$O → (~R □→ O)

So, (7) ~$◊_m$O [from (5) and (6)]

for saving his grandfather's life, it is clear that this possible scenario constitutes a counterexample to (PAP). I take this case against (PAP) to be fairly strong.

The discussion above may seem familiar to those who are acquainted with Kadri Vihvelin's (1996) work on time travel. Vihvelin and I agree that time travel is possible and that there are certain limitations on what a time traveler can or could have done. Vihvelin argues that a time traveler could not kill her younger self under certain circumstances and I argue that Martin could not do otherwise than release the emergency safety net under certain circumstances (and that he is nevertheless responsible for doing so). Our arguments are slightly different from one another, though. Vihvelin's main premise is that if a time traveler could have killed her younger self, then she would or at least might have succeed in doing so if she had tried. As Peter Vranas (2010) has recently pointed out, there are some things we can do but would fail to do if we tried. To borrow an example from Vranas, I can type 90 words per minute without trying, but if I were to try, I would become self-conscious and fail to do so. Some might take cases like this one as a reason to doubt that Vihvelin's main premise is true. On the other hand, my related premise is that if Martin could have done otherwise than release the emergency safety net, then he would have been doing otherwise if he had not released the safety net. This premise is not subject to the kind of worry Vranas presents because it doesn't involve the notion of *trying* at all.[9] Of course, the most popular objection to Vihvelin's argument involves an appeal to the context sensitivity

---

The inference from (3) and (4) to (5) is needed to rule out that (3) is vacuously true. If (3) is vacuously true, then (5) is the denial of a vacuous truth and, hence, clearly false. But, of course, if (4) is true, then (3) is not vacuous and (5) is true as well.

[9] Vranas does present other objections in his paper as well. Unfortunately, it would be too much of a detour to discuss his other objections here.

of 'can'.[10] Such an objection may have some force against my argument as well. So, I will consider this objection in more detail below.

My counterexample might also seem strikingly similar to Fischer-Style Counterexamples to principles like (PAP).[11] According to one Fischer-Style Counterexample, a man who is going to carry out a particularly heinous act for which he will be morally responsible is being watched over by an assassin who would kill the man if the man were to decide to refrain from performing the act in question. For example, imagine that Steve has made a deal with the mob and agreed to murder a particular person on Saturday morning. The mob, though, in order to ensure the cooperation of people who make deals with them in the future, plans on killing Steve before Saturday morning if he shows any sign of hesitation, any sign that he will back out of the deal. Steve, being a trustworthy person, does not hesitate and does kill the person he promised to kill. According to Fischer-Style Counterexamples, Steve cannot do otherwise than kill the man he has promised to kill on Saturday morning. This is because if Steve had hesitated, then he would have been killed before Saturday morning and so would not have even existed and, hence, would not have *done* anything that morning; in particular, he would not have done otherwise than kill the man he had promised to kill. So, Steve could not have done otherwise than kill the man he promised to kill. However, it is clear that Steve is morally responsible for that killing. So, according to this Fischer-Style Counterexample, (PAP) is false.

There are some similarities between this Fischer-Style Counterexample and the time travel counterexample introduced in this paper. In both cases, there is a person who will not exist

[10] I take this to be the core of Sider's (2002) objection and of Ira Kiourti's (2008) objection as well.

[11] Fischer Style Counterexamples were discussed by John Martin Fischer (2003)

in various counterfactual scenarios. Moreover, because the person fails to exist in those counterfactual scenarios, he is, putatively, not able to do otherwise than what he in fact does.

However, even though these cases are similar in certain respects, there are also important dissimilarities between them. One dissimilarity that is very important is that in the Fischer-Style Counterexample, the individual who is morally responsible yet, putatively, could not do otherwise, does exist in the counterfactual scenarios at some time. In our example above, the man who has made a deal with the mob, Steve, would have existed if he had hesitated and been killed by the mob; he just wouldn't have existed on Saturday morning. But, a plausible objection to Fischer-Style Counterexamples is that a person need not have existed at a particular time in order to do otherwise than a particular action at that time. If I know that I will be forced to do some particularly heinous act and I know that the only way I can avoid doing that act is by ensuring that I am dead before Saturday, then if I were to kill myself, it seems clear that I would in fact have done otherwise than that particular heinous act.[12]

In the new time travel counterexample, Martin, the man who saves his grandfather, would not have existed *at all*, at any time, if he had not released the emergency safety net. The plausible objection to the Fischer-Style Counterexample introduced above relied on the fact that Steve would have existed at some time even if he had hesitated and been killed by the mob before Saturday. Since Martin would not have existed at any time at all if he had not released the emergency safety net, it is clear that there is no analogous plausible objection to the time travel counterexample. So, it is clear that, although Fischer-Style Counterexamples and the new time travel counterexample are similar in certain respects, the new time travel counterexample is a genuinely new case and it has certain advantages over Fischer-Style Counterexamples.

[12] This point is also made by in Maria Alvarez (2009).

Another interesting thing to note about the new time travel counterexample is that, just as it has certain advantages over Fischer-Style Counterexamples, it also has advantages over the original Frankfurt-Style Counterexamples. Remember that one plausible objection to the Frankfurt-Style Counterexample introduced above involves claiming that Jones in fact could have done otherwise than kill Smith by ensuring that his bodily movements don't constitute an act of killing. This objection relies on the claim that Jones's bodily movements would not constitute an act of killing if they were controlled by the mechanism implanted in his brain. But, this objection also relies on the claim that Jones would have existed if he had chosen to do otherwise than kill Smith. After all, it is impossible for Jones to do otherwise than kill Smith if he does not exist at all. However, in the time travel counterexample, Martin would not exist if he had not released the emergency safety net. Since Martin doesn't exist in the relevant counterfactual scenarios, he can't do anything at all. The distinction between genuine actions and mere bodily movements doesn't help to undermine the time travel counterexample. So, it seems that the new time travel counterexample has an advantage over the original Frankfurt-Style counterexamples.

In the remainder of this essay, I consider and respond to some plausible objections to the new time travel counterexample to (PAP). After finding those objections unpersuasive, I show how the time travel objection can be modified as an objection to principles that some might take to be immune to the problems faced by (PAP).

3. Objections to the first premise

The counterexample to (PAP) relies on the argument for the claim that Martin could not have done otherwise than save his grandfather. So, any objection to that argument will constitute

a strong defense of (PAP). In this section I will consider two defenses of (PAP) each one of which involves objecting to the first premise of the argument above.

First, one might object to this argument and defend (PAP) by pointing out that Martin's grandfather might have survived the fall even if Martin had not pushed the button to release the emergency safety net. If, for example, another person had been in the area and had pushed the button to release the safety net, then Martin's grandfather would have survived even if Martin himself refrained from pushing the button; if Martin's grandfather had seemingly miraculously bounced without being harmed by the fall, then he would have survived even if Martin had refrained from pushing the button; if Martin's grandfather had been rescued by the tractor beam of some benevolent aliens or by the powerful will of some benevolent deity, then he would have survived even if Martin had refrained from pushing the button. These are just some of the many possible situations in which Martin's grandfather survives his fall while Martin refrains from pushing the button. So, one might think that the first premise in my argument above, the premise that says if Martin had not released the emergency safety net, then Martin would have never existed, is false.

I hope you'll agree that this objection is off the mark. Sure it's possible for someone else to be near the button when Martin's grandfather falls from his high-wire. But, in the case we described, Martin and his grandfather are in a desolate area and no one else is around. Sure it's possible for Martin's grandfather to bounce without being harmed by the fall. But, this is highly unlikely and, we might even suppose, physically impossible given the grandfather's trajectory. It's possible for aliens or a benevolent deity to save Martin's grandfather. But, let's simply stipulate that there are no aliens or deities, or that any such creatures are obeying a prime

directive of non-interference.[13] In other words, let's just stipulate that absolutely no one else is in a position to save Martin's grandfather and that there is absolutely no way for Martin's grandfather to survive his fall without being rescued by Martin. This is a genuine possibility. Moreover, if Martin and his grandfather were in such a bleak circumstance, then it wouldn't matter if there were distant possibilities in which Martin's grandfather survives even though Martin does not release the safety net. These distant possibilities do not act as guardian angels; they do not protect truths about what Martin can do.

A second objection involves claiming that the first premise presupposes the necessity of origins. That is, the objector claims, the first premise presupposes that Martin essentially has the grandfather he in fact has. But, if Martin could have had a different grandfather than he in fact has, then it might not matter to his survival whether he saves his grandfather or not. So, if the presupposition is mistaken, then perhaps Martin can exist even if his grandfather dies in the fall. That is, perhaps premise (1) is false.

It's absolutely right that if the necessity of origins is true, then the first premise is guaranteed to be true. However, the first premise is true even if the necessity of origins is false. All that is needed to make the premise true is that Martin counterfactually depends for his existence on his grandfather; if Martin's grandfather had not existed, then he would not have existed either. But, this seems right. Even if the necessity of origins is false, the possibilities in which Martin fails to have the grandfather he in fact has are quite distant. Even the most plausible possibilities are still distant possibilities. Perhaps amongst the plausible possibilities are

[13] Some Christian theists seem to believe that God obeys a prime directive of non-interference. Some believe, for example, that such a directive must be followed in order for humans to have free will and in order for the world to have whatever good results from the fact that humans have free will.

ones in which Martin is genetically engineered by a technologically advanced race, or ones in which Martin was born by some genetic fluke from completely different parents, or ones in which Martin's grandfather had a twin, or ones in which God creates Martin *ex nihilo*. But these distant possibilities are not the nearest ones in which Martin fails to release the emergency safety net. These distant possibilities cannot undermine premise (1). The nearest possibilities in which Martin fails to release the emergency safety net are ones in which his grandfather dies and in which he never exists.

So it seems that neither of these objections to the first premise of my argument is sound. And so, neither objection can be used in a successful defense of (PAP).

4. Grandfather Paradoxes and Contextualism

In some respects, the time travel counterexample to (PAP) is similar to the classic grandfather paradox. According to the grandfather paradox, if it's possible for a man, Tim, to travel into the past and meet his young grandfather, then such a time traveler can kill his own grandfather before his grandfather has any children or in any way preserves his gametes for posterity. But, since the grandfather's death at such a time would ensure that the time traveler never exists, no time traveler can kill his own grandfather in such circumstances. Thus, if it is possible for a man to travel into the past and meet his young grandfather, then such a time traveler both can and cannot kill his grandfather.

One might think that a similarly paradoxical situation arises with Martin. If it's possible a time traveler save his young grandfather before the grandfather has any children or in any way preserves his gametes for posterity, then such a time traveler can also do otherwise than save his grandfather in such circumstances. However, since the grandfather's death at such a time would

ensure that the time traveler never exists, it also seems that no time traveler can do otherwise than save his grandfather in such circumstances. So, a time traveler who saves his grandfather both can and cannot do otherwise than save his grandfather.

David Lewis has famously argued that one lesson we should learn from the time travel paradoxes is that phrases of the form 'S can do A' are context sensitive. Lewis's account of such phrases can be stated as follows:

(LAC) Any English sentence of the form 'S can do A' expresses a truth in context C iff S's doing A is consistent with the facts that are salient in C.[14]

According to Lewis, in some contexts, the sentence 'Tim can kill his grandfather' expresses a truth; it expresses a truth in those contexts where the salient facts include facts about Tim's physiology and do not include facts about Tim's origins or time traveling adventures. However, in other contexts 'Tim can kill his grandfather' expresses a falsehood. Those are contexts in which facts about Tim's origins and his time traveling adventures are salient. Lewis claims that we are faced with a seemingly paradoxical situation when we consider these time travel scenarios because we fail to recognize a subtle shift in context.

[14] Lewis uses the word 'compossible' rather than 'consistent'. If 'P is compossible with Q' means that P and Q are together possible, then it may be that Lewis's phrasing, along with the claim that 'Tim can kill his grandfather in the circumstances in question' expresses a truth in some contexts, will commit him to a denial of the necessity of origins or to an acceptance of the possibility of branching time or resurrection. Otherwise, it would be impossible for Tim to kill his grandfather, and hence not compossible with anything. Using the word 'consistent' instead will result in no such commitment.

We might think, similarly, that 'Martin can do otherwise than release the emergency safety net' expresses a truth in some contexts and a falsehood in others. It expresses a truth in those contexts where the salient facts include facts about Martin's physiology and do not include facts about Martin's origins or time traveling adventures. On the other hand, that same sentence expresses a falsehood in contexts in which facts about Martin's origins and time traveling adventures are salient. Again, we are faced with a seemingly paradoxical situation because we fail to recognize a subtle shift in context.

If Lewis's response to the grandfather paradox is right, then it seems like there should be a plausible defense of (PAP) from the time travel objection introduced above. This is because in some contexts we would be speaking truly if we were to say "Martin can do otherwise than release the emergency safety net" and in others we would not. A defender of (PAP) should say that the sentences beside the numbers (1)-(7) above express a sound argument in some contexts and do not express a sound argument in other contexts.[15]

Although I believe a contextualist defense of (PAP) would be quite formidable, I do not believe it would ultimately be successful. I do not want to spend too much time presenting and evaluating a contextualist defense of (PAP), but I will briefly present an argument against one kind of contextualist position.

[15] Most likely such a defender should say that the sentence next (1) expresses a falsehood in some contexts. Of course, such a response will most likely commit the defender to the claim that counterfactuals are context sensitive. Although the defender of (PAP) may not be worried about such a commitment (since it is widely accepted that counterfactuals are context sensitive) she should provide a theory that indicated the connection between those contexts in which the counterfactual is false and those contexts in moral responsibility attributions are correct.

Consider a contextualist position according to which the sentence I used to express (PAP), I'll refer to it with the description 'the (PAP) sentence', expresses a truth in every context if it expresses a truth at all. I will show that this kind of contextualist position is mistaken. If the (PAP) sentence expresses a truth in every context, then the sentence 'Martin is morally responsible for releasing the safety net only if Martin could have done otherwise than release the safety net' also expresses a truth in every context. After all, this particular sentence about Martin is just a logical consequence of an instance of the (PAP) sentence. But, according to the contextualist defense of (PAP), 'Martin could have done otherwise than release the safety net' expresses a falsehood in some contexts. Since 'Martin is morally responsible for releasing the safety net only if Martin could have done otherwise than release the safety net' expresses a truth in every context and 'Martin could have done otherwise than release the safety net' expresses a falsehood in some contexts, it follows that 'Martin is morally responsible for releasing the safety net' expresses a falsehood in some contexts as well.[16] It seems clear, though, that that sentence also expresses a truth in some contexts. But, if 'Martin is morally responsible for releasing the safety net' expresses a truth in some contexts and a falsehood in others, then it is a context sensitive sentence. So, if the (PAP) sentence expresses a truth in every context, then "Martin is morally responsible for releasing the safety net" is context sensitive.

Perhaps some are willing to accept that sentences like "Martin is morally responsible for releasing the safety net" are context sensitive, but I am not.[17] It seems that if "Martin is morally

[16] I am assuming here that the logical truth conditions for conditional sentences remains fixed across contexts.

[17] Admittedly, whether or not someone is morally responsible for an action depends on the details of the circumstances under which the action is performed. In some circumstances, one may be morally responsible for stealing and in others one may not. But, this view is a view about being morally responsible not about the words

responsible for releasing the safety net" expresses a truth in my context, then (all else being equal) either I act appropriately if I praise Martin or I act appropriately if I blame Martin for his action. It might, for example, be appropriate for me to congratulate Martin or it might be appropriate for me to punish Martin. However, if "Martin is morally responsible for releasing the safety net" expresses a falsehood in your context, then (all else being equal) you act inappropriately if you either praise or blame Martin for his action; it is not appropriate for you to congratulate Martin or to punish Martin. But, it hardly seems fair or just to Martin if it is appropriate for people to treat him differently solely in virtue of facts about *their* conversational contexts. So, it seems that "Martin is morally responsible for releasing the safety net" is not context sensitive.[18] I take this to be a sufficient reason to reject any view on which the (PAP) sentence expresses a truth in every context and yet sentences of the form 'S could have done A' do not.

Since the contextualist defense of (PAP) essentially involves the claim that sentences of the form 'S could have done A' are context sensitive, it looks like the contextualist defender of (PAP) must deny that the (PAP) sentence expresses a truth in all contexts. I will not consider this alternative here. But, I will mention that such a contextualist must come up with a plausible theory about those conversational contexts in which the (PAP) sentence expresses a truth. The

---

"morally responsible". Hence, it does not imply contextualism. If a person is morally responsible for stealing in a particular circumstance, then it doesn't matter what conversational context we are in, if we are talking about the person's moral responsibility, we speak truly when we say "she is morally responsible for stealing" and falsely when we say "she is not morally responsible for stealing".

[18] I am very grateful to Kelly McCormick for suggesting and extensively discussing this argument with me. Unfortunately, this argument deserves a lengthier discussion than is warranted for this paper.

challenge to the contextualist is to come up with some morally relevant feature of certain conversational contexts that grounds both the truth of the (PAP) sentence and the falsity of one of sentences (1)-(7).

5. Extending the Counterexample

When I first introduced (PAP) above, I said that the quantifiers in (PAP) range over both actions that are fully determinate and those that are not. One might think, however, that fully determinate actions play an important role in moral philosophy. Perhaps a variant of (PAP) that makes reference to fully determinate actions will show deference to the important role that such actions play in moral philosophy while generating a principle that is immune to the various styles of counterexamples that we have been considering. With this in mind, we can formulate a principle, which I will call 'The Sophisticated Principle of Alternate Possibilities' or '(SPAP)', as follows:

(SPAP) Necessarily: for any person S and any fully determinate action A, S is morally responsible for performing A only if there is some fully determinate action A* such that S could have done A* while failing to do A.

Remember that a fully determinate action cannot be performed in multiple ways. Let's take this claim to be very strong. Let's take this claim to be strong enough to entail that if I had typed the word at the beginning of this sentence with a slightly different amount of pressure on the keys or if my fingers had even been positioned a little differently, then I would have performed a different fully determinate action than I in fact perform. Let's even suppose that if I had

performed an action with slightly different causal origins than the action I in fact performed, then I would have performed a different fully determinate action than the one I actually performed. Taken in this way, the principle (SPAP) is very weak.[19]

This principle is not subject to Frankfurt-Style Counterexamples. Even if your bodily motions, controlled by the mechanism in your brain, are actions (contrary to what I suggested earlier), Frankfurt-Style Counterexamples leave (SPAP) unscathed. Consider Jones, the man who wants to murder Smith. If he had decided to do otherwise than murder Smith, then the mechanism that Black implanted in his brain would have activated and Jones would have done a different fully determinate action than he in fact did; the fully determinate action he actually performed was not caused at all by the mechanism in his head whereas the fully determinate action he would have done if he had tried to do otherwise would have been partially caused by the mechanism in his head. (SPAP) is safe from Frankfurt-Style Counterexamples.

What about the new time travel counterexample? Martin's adventures into the past show that (PAP) is false, but do they also show that (SPAP) is false? Initially, one might think that (SPAP) is immune to time travel counterexamples. Consider the case of Martin and his grandfather. Even though Martin could not have done otherwise than release the emergency safety net, he could have done otherwise than release the safety net in the particular way he did. Martin could have pressed the button with a slightly greater force than he actually did; Martin could have pressed the button with his finger in a slightly different position than it was actually

[19] Peter van Inwagen (1983) considers a principle that involves act particulars, each one of which has its causal origins essentially. Insofar as I understand act particulars, they are simply fully determinate actions. John Martin Fischer (2006:13) suggests that it is impossible to come up with a counterexample to principles that employ such fine grained notions of action. In what follows I hope to show that Fischer is mistaken about this.

in; Martin could have pressed the button while moving fluidly or he could have done so while moving rigidly. In short, Martin could have performed a different fully determinate action than he in fact performed. So, it seems that the new time travel counterexample leaves (SPAP) unscathed as well.

However, there are a number of ways we can fill out the time travel case. Let's suppose that Martin first found his time machine abandoned in a field. The machine had been built by a race of extremely powerful beings who constructed a rather intricate locking mechanism in the device. The locking mechanism prevented the machine from being activated unless a particular fully determinate event occurred at some time in the past. That fully determinate event was the event of Martin's pushing the button to release the emergency safety net in the particular way he did. If, in the past, Martin had not pushed the button in the particular way he did, if, in other words, he had not performed the particular fully determinate action that he actually performed, then the time travel machine would have remained locked forever, no one would have been able to use it to travel in time, Martin's grandfather would have plummeted to his death in the desolate wilderness and Martin would never have existed. But, of course, if Martin had not existed, then he would not have been able to *do* any fully determinate action; in particular, he would not have been able to do any fully determinate action different from his actual action. So, it seems that it doesn't matter whether or not the actions in question are fully determinate. We can construct a time travel counterexample to (SPAP) simply by adding a few details to our time travel story.

6. Conclusion

In this paper, I introduced a new kind of counterexample to the Principle of Alternative Possibilities. This new counterexample is distinct from both Frankfurt-Style Counterexamples and Fischer-Style Counterexamples; this new kind of counterexample has advantages over both of them. Moreover, even (SPAP), a principle that appeals to fully determinate actions rather than actions that can be performed in multiple ways, can be shown to be false with a time travel counterexample. It seems, then, that the future looks bleak for principles of alternative possibilities.[20]

Bibliography

Alvarez, M. (2009). "Actions, Thought-Experiments and the 'Principle of Alternate Possibilites'". *Australasian Journal of Philosophy* 87: 61-81.

Arntzenius, F. and Maudlin, T. (2002). "Time travel in modern physics" in *Time, Reality and Experience*, ed. C. Callender. Cambridge University Press, Cambridge.

Bricker, P. (1991). "Plenitude of Possible Structures" *The Journal of Philosophy* 88: 607-619.

Davies, P. (2001). *How to Build a Time Machine*. Viking Penguin: New York.

Fischer, J. M. (2003). "Responsibility and Agent Causation" in *Moral responsibility and Alternative Possibilities,* ed. David Widerker and Michael McKenna, Aldershot, UK: Ashgate.

Fischer, J. M. (2006) "Introduction" in *My Way*. Oxford University Press: New York.

[20] Thanks to Andrew Cullison, Ben Caplan, Hud Hudson, André Gallois, Kelly McCormick, Neal Tognazzini, and an anonymous referee for several helpful comments and suggestions. Thanks to the anonymous referee for suggesting the title for this paper. Earlier drafts of this paper were presented at University of Wisconsin - Milwaukee and at SUNY Fredonia (for the Young Philosophers lecture series). Thanks to both audiences.

Frankfurt, H. (1969). “Alternate Possibilities and Moral Responsibility”. *Journal of Philosophy* 66: 829-39.

Kiourti, I. (2008). “Killing Baby Suzy”. *Philosophical Studies* 139: 343-352.

Lewis, D. (1976). “The Paradoxes of Time Travel” *American Philosophical Quarterly* 13, pp 145-152. Reprinted in *Philosophical Papers Vol. II*. Oxford: Oxford University Press: New York. pp 67-80.

Sider, T. (2002). “Time Travel, Coincidences and Counterfactuals”. *Philosophical Studies* 110: 115-138.

van Inwagen. P. (1983). *An Essay on Free Will.* Oxford University Press: New York.

Vihvelin, K. (1996). “What Time Travelers Cannot Do”. *Philosophical Studies* 81: 315-330.

Vranas P. B. M. (2010). “What time travlers may be able to do”. *Philosophical Studies* 150:115-121.